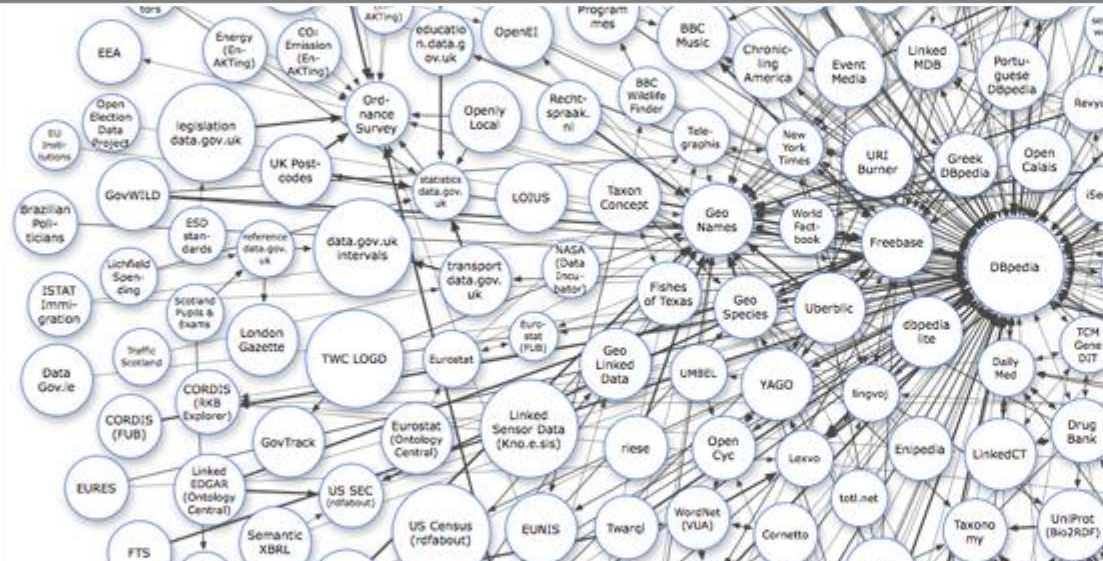


# Applying Linked Data Technologies to Financial Reporting

**Benedikt Kämpgen, André Freitas, Andreas Harth, Seán O’Riain**

# Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web, October 16, 2012

Institute of Applied Informatics and Formal Description Methods (AIFB)



# Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- UC2: Understanding the Origin and Trustworthiness of XBRL
- Relation to DDI Lifecycle

# Financial analysis



Thanks to Craig Weber for this example

**Task:** Find sales and operating income for each segment of company 三精輸送機

# Financial analysis (2)



EDINET - Mozilla Firefox

info.edinet-fsa.go.jp

ださい。

**お知らせ**  
○EDINETの受理時間について  
株券の発行において、当日の市場価格の終値を参照して発行価格が確定し、当日又は翌日に効力を生じさせることが必要な訂正届出書について、発行価格の協議に時間を要する場合、書類作成に時間を要する場合、システムの都合等やむを得ない理由があるときは、原則として、EDINETの受付時間を延長することになりました（おおむね19時まで）。※なお、その他の書類の開示についても、EDINETの受付時間を延長することがあります。

**更新情報**  
平成24年10月4日 [次世代EDINETタクソノミ\(案\)第二版並びに提出者向け事前チェックテスト実施要領及び参加要領](#)が公表されました。  
平成24年10月4日 [提出書類ファイル仕様書](#)が更新されました。  
平成24年6月29日 [よくある質問](#)が更新されました。(情報公開)  
平成24年5月25日 [企業別タクソノミ作成ガイドライン\(その2:IFRS適用提出者用\)](#)が更新されました。  
平成24年4月11日 [推奨端末仕様](#)が更新されました。  
平成24年3月14日 [2012年版EDINETタクソノミ及びXBRL作成ガイド](#)が公開されました。  
平成24年3月14日 [EDINET概要書](#)が更新されました。  
平成24年1月28日 [大量保有報告書様式\(Excel版\)の使い方について](#)が更新されました。  
平成23年7月25日 [2011年版EDINETタクソノミの追加タクソノミ](#)が公開されました。  
平成23年3月14日 [2011年版EDINETタクソノミ及びXBRL作成ガイド](#)が公開されました。

**閲覧**  
有価証券報告書等 公告

**ダウンロード**  
→ [XBRL一括](#) → [EDINETタクソノミ](#) → [EDINETコードリスト](#)  
→ [提出書類様式](#) → [EDINETコード対応表](#)

金融庁/Financial Services Agency, The Japanese Government Copyright©金融庁 All Rights Reserved.

# Financial analysis (2)



**Extensible Business Reporting Language (XBRL):** Describes business reporting information in an XML format, for exchange, and comparison [XBRL SPEC].



# XBRL – Example



**View Filing Data**

SEC Home » Search the Next-Generation EDGAR System » Company Search » Current Page

**RAYONIER INC (Filer) CIK: 0000052827**

Print Document View Excel Document

Cover	CONDENSED CONSOLIDATED STATEMENTS OF INCOME AND COMPREHENSIVE INCOME (USD \$) In Thousands, except Per Share data		3 Months Ended		9 Months Ended	
	Sep. 30, 2010	Sep. 30, 2009	Sep. 30, 2010	Sep. 30, 2009	Sep. 30, 2010	Sep. 30, 2009
Document and Entity Information						
Financial Statements						
CONDENSED CONSOLIDATED STATEMENTS OF INCOME AND COMPREHENSIVE INCOME						
CONDENSED CONSOLIDATED STATEMENTS OF INCOME AND COMPREHENSIVE INCOME (Parenthetical)						
CONDENSED CONSOLIDATED BALANCE SHEETS						
CONDENSED CONSOLIDATED BALANCE SHEETS (Parenthetical)						
CONDENSED CONSOLIDATED STATEMENTS OF CASH FLOWS						
Notes to Financial Statements						
	SALES	\$ 377,515	\$ 300,648	\$ 999,925	\$ 858,731	
	<b>Costs and Expenses</b>					
	Cost of sales	269,203	231,836	744,996	672,855	
	Selling and general expenses	17,125	15,972	49,264	44,962	
	Other operating income, net (Note 2)	(792)	(59,251)	(6,620)	(150,425)	
	Costs and Expenses, Total	285,536	188,557	787,640	567,392	
	Equity in income (loss) of New Zealand joint venture	103	(943)	634	(2,782)	
	OPERATING INCOME BEFORE GAIN ON SALE OF A PORTION OF THE INTEREST IN THE NEW ZEALAND JOINT VENTURE	92,082	111,148	212,919	288,557	
	Gain on sale of a portion of the interest in the New Zealand joint venture (Note 6)			12,367		
	OPERATING INCOME	92,082	111,148	225,286	288,557	
	Interest expense	(12,943)	(12,789)	(37,680)	(37,630)	
	Interest and miscellaneous income, net	345	310	943	594	
	INCOME BEFORE INCOME TAXES	79,484	98,669	188,549	251,521	
	Income tax expense	(16,580)	(17,529)	(30,134)	(36,707)	
	NET INCOME	62,904	81,140	158,415	214,814	
	<b>OTHER COMPREHENSIVE INCOME (LOSS)</b>					
	Foreign currency translation adjustments	3,198	2,620	(64)	13,568	
	Joint venture cash flow hedges	(104)	968	922	(1,659)	
	Amortization of pension and					

- U.S. Securities and Exchange Commission database
- RAYONIER INC had a sales revenue net of 377,515,000 USD from 2010-07-01 to 2010-09-30 [XBRL example]

# XBRL – Example



**View Filing Data**

SEC Home » Search the Next-Generation EDGAR System » Company Search » Current Page

**RAYONIER INC (Filer) CIK: 0000052827**

Print Document View Excel Document

Cover	CONDENSED CONSOLIDATED STATEMENTS OF INCOME AND COMPREHENSIVE INCOME (USD \$) In Thousands, except Per Share data			
	3 Months Ended		9 Months Ended	
Document and Entity Information	Sep. 30, 2010	Sep. 30, 2009	Sep. 30, 2010	Sep. 30, 2009
Financial Statements				
CONDENSED CONSOLIDATED STATEMENTS OF INCOME AND COMPREHENSIVE INCOME				
SALES	\$ 377,515	\$ 300,648	\$ 999,925	\$ 858,731
Costs and Expenses				
Cost of sales	269,203	231,836	744,996	672,855
Selling and general expenses	17,125	15,972	49,264	44,962
Other operating income, net (Note 2)	(792)	(59,251)	(6,620)	(150,425)
Costs and Expenses, Total	285,536	188,557	787,640	567,392
Equity in income (loss) of New Zealand joint venture	103	(943)	634	(2,782)
OPERATING INCOME BEFORE GAIN ON SALE OF A PORTION OF THE INTEREST IN THE NEW ZEALAND JOINT VENTURE	92,082	111,148	212,919	288,557
Gain on sale of a portion of the interest in the New Zealand joint venture (Note 6)			12,367	
OPERATING INCOME	92,082	111,148	225,286	288,557
Interest expense	(12,943)	(12,789)	(37,680)	(37,630)
Interest and miscellaneous income, net	345	310	943	594
INCOME BEFORE INCOME TAXES	79,484	98,669	188,549	251,521
Income tax expense	(16,580)	(17,529)	(30,134)	(36,707)
NET INCOME	62,904	81,140	158,415	214,814
OTHER COMPREHENSIVE INCOME (LOSS)				
Foreign currency translation adjustments	3,198	2,620	(64)	13,568
Joint venture cash flow hedges	(104)	968	922	(1,659)
Amortization of pension and				
Notes to Financial Statements				

- U.S. Securities and Exchange Commission database
- RAYONIER INC had a sales revenue net of 377,515,000 USD from 2010-07-01 to 2010-09-30 [XBRL example]

# Outline

- Extensible Business Reporting Language (XBRL)
- **UC1: Integrating XBRL with External Data Sources**
- UC2: Understanding the Origin and Trustworthiness of XBRL
- Relation to DDI Lifecycle



# UC1: Integrating XBRL with External Data Sources

## ■ CalcBench: Comparison of heterogeneous reports

RAYONIER INC [VIEW IN SPREADSHEET](#)

Condensed Consolidated Statements of Income and Comprehensive Income - Quarterly → [Switch to Yearly View](#)

[EXPORT](#) (log in to post comments, save data)

Link here: [www.calcbench.com/t/RYN/i/q](http://www.calcbench.com/t/RYN/i/q)

Compare RYN to:

2Q 2012  
4/1 - 6/30/2012

COMPARE TO COMPANY:  
WELLPOINT, INC  
2Q 2012  
4/1 - 6/30/2012

[hide](#)

SALES	\$371,926,000	\$3,314,200,000	(88.78%)
Costs and Expenses			
Cost of sales	\$262,555,000	\$6,700,000	3,818.73%
Name: Adj. Cost of Sales X share	262,000,000	6,500,000	
Selling and general expenses	\$16,250,000	\$2,081,800,000	(99.22%)
Other operating (income) expense, net	(\$5,299,000)	-	
Costs and Expenses, Total	\$273,506,000	-	
Equity in income of New Zealand joint venture	\$170,000	\$0	
OPERATING INCOME	\$98,590,000	\$14,358,700,000	(99.31%)
Operating Margin %	26.51%	433.25%	
Interest expense	(\$16,056,000)	\$117,600,000	(86.35%)
Interest and miscellaneous income, net	\$85,000	\$1,048,600,000	(99.99%)
INCOME BEFORE INCOME TAXES	\$82,619,000	\$1,048,600,000	(92.12%)
Income tax expense	(\$13,540,000)	\$405,000,000	(96.66%)
NET INCOME	\$69,079,000	\$643,600,000	(89.27%)
Net Income Margin %	18.57%	19.42%	

# UC1: Integrating XBRL with External Data Sources

## ■ CalcBench: Comparison of heterogeneous reports

RAYONIER INC [VIEW IN SPREADSHEET](#)

Condensed Consolidated Statements of Income and Comprehensive Income - Quarterly → [Switch to Yearly View](#)

[EXPORT](#) (log in to post comments, save data)

Link here: [www.calcbench.com/t/RYN/i/q](http://www.calcbench.com/t/RYN/i/q)

Compare RYN to:

2Q 2012  
4/1 - 6/30/2012

COMPARE TO COMPANY:  
WELLPOINT, INC  
2  
/30/2012

Computation of missing values  
(e.g., total revenues)

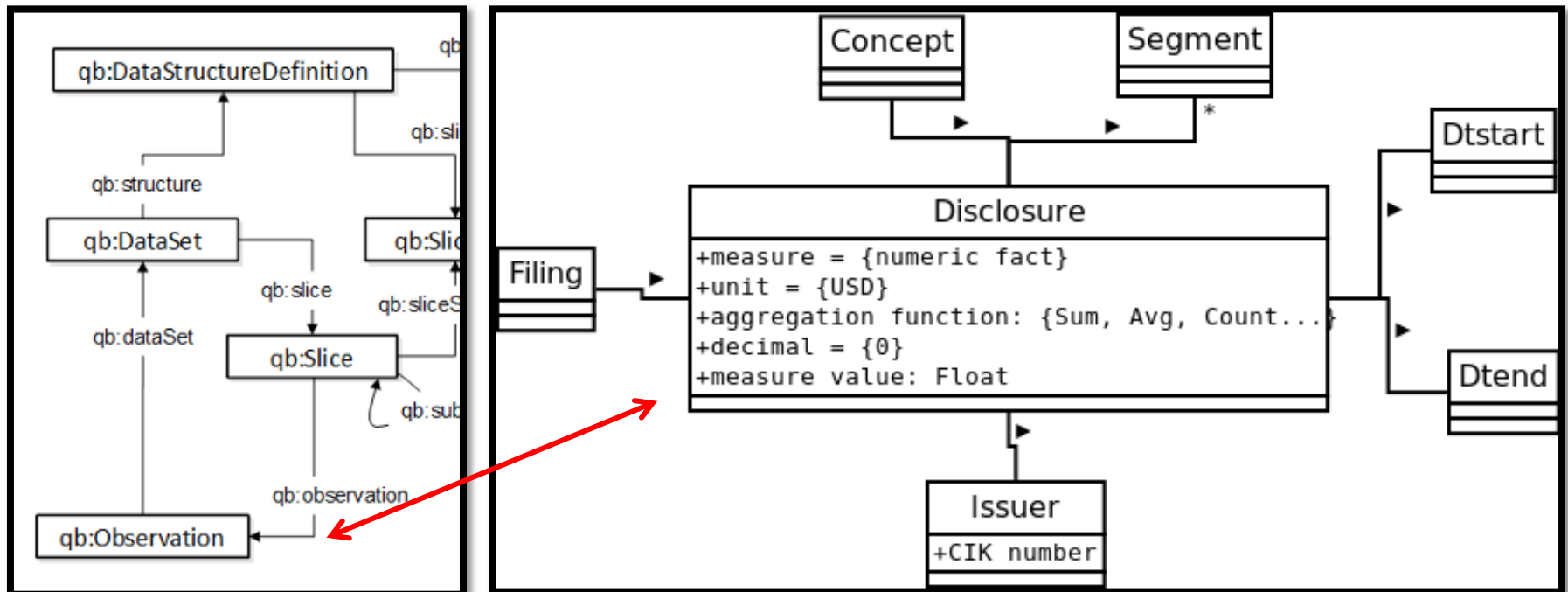
SALES	\$371,926,000	\$3,314,200,000	(81.78%)
Costs and Expenses			
Cost of sales	\$262,555,000	\$6,700,000	3,818.73%
Name: Adj. Cost of Sales X share	262,000,000	6,500,000	
Selling and general expenses	\$16,250,000	\$2,081,800,000	(99.22%)
Other operating (income) expense, net	(\$5,299,000)	-	
Costs and Expenses, Total	\$273,506,000	-	
Equity in income of New Zealand joint venture	\$170,000	\$0	
OPERATING INCOME	\$98,590,000	\$14,358,700,000	(99.31%)
Operating Margin %	26.51%	433.25%	
Interest expense	(\$16,056,000)	\$117,600,000	(86.35%)
Interest and miscellaneous income, net	\$85,000	\$1,048,600,000	(99.99%)
INCOME BEFORE INCOME TAXES	\$82,619,000	\$1,048,600,000	(92.12%)
Income tax expense	(\$13,540,000)	\$405,000,000	(96.66%)
NET INCOME	\$69,079,000	\$643,600,000	(89.27%)
Net Income Margin %	18.57%	19.42%	

# UC1: Scenarios

- Can we group SEC financial companies by **operation industry** from **Freebase** and get the sales revenue net and cost of goods sold to compute the average **Gross Profit Margin**?
- Can we compare the **sales revenue net** and **cost of goods sold** of SEC companies with expenses for **automobile loans** reported in FFIEC call reports?

# UC1: Approach

- Using RDF Data Cube Vocabulary (QB) for representing XBRL data



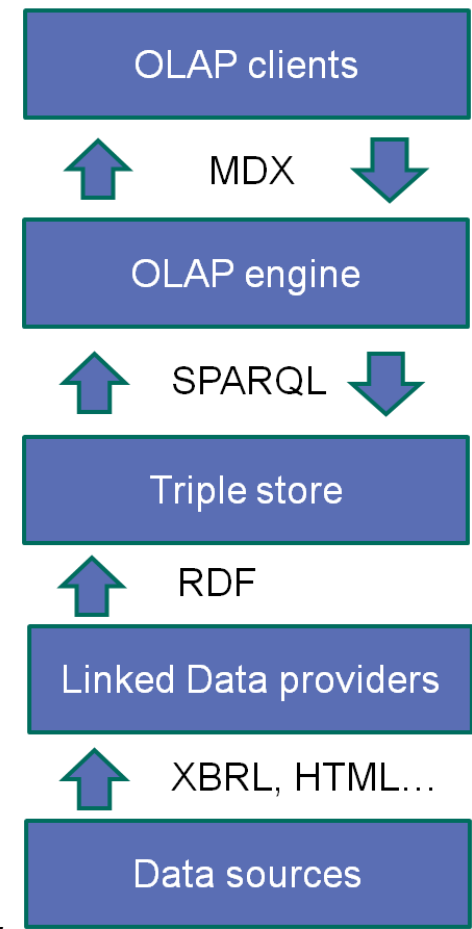
[QB] <http://www.w3.org/TR/vocab-data-cube/>

# UC1: Approach – Architecture

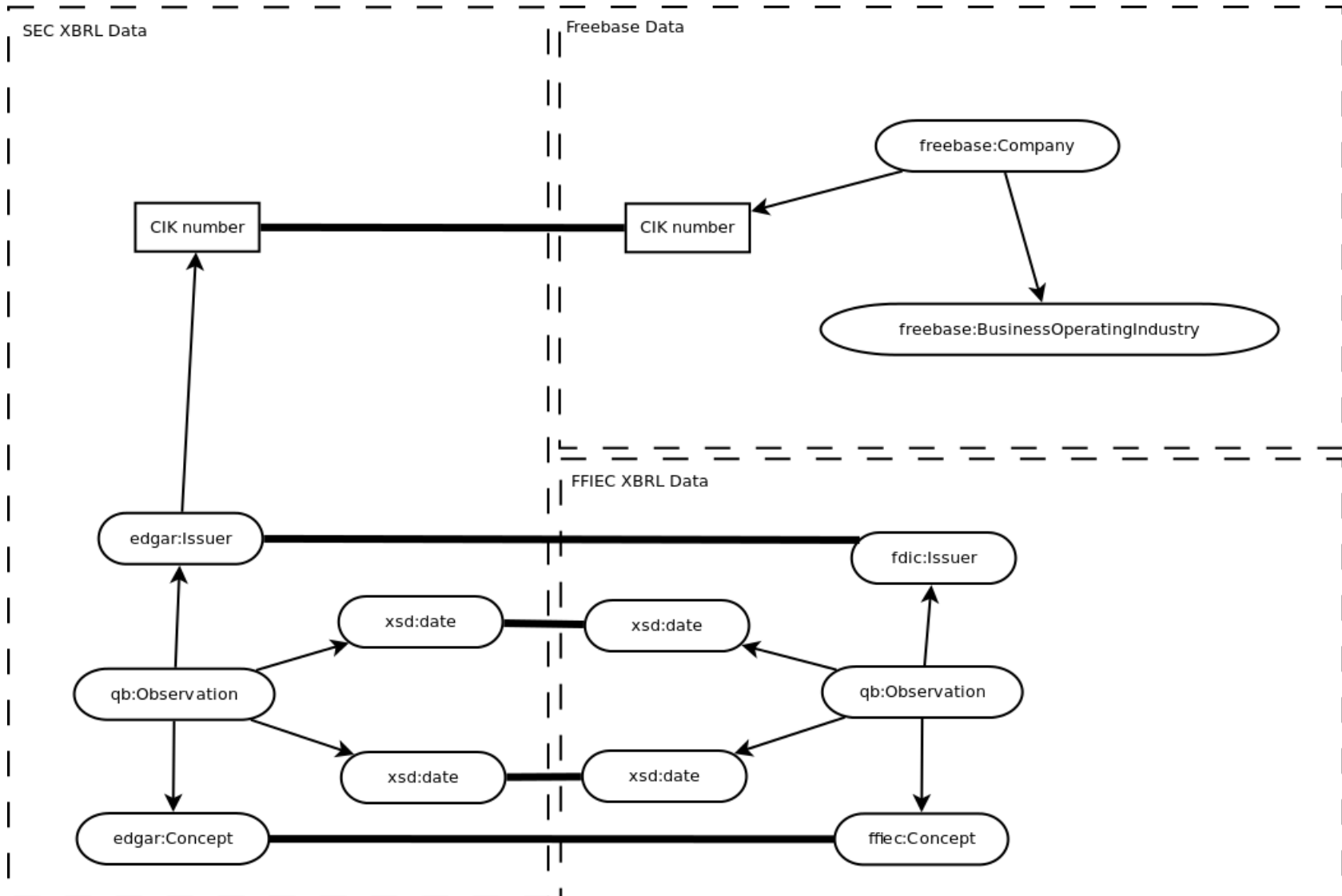
- Saiku OLAP client
- Olap4Ld
- Open Virtuoso
- LDSpider crawling
  - SEC XBRL database (Edgar Linked Data Wrapper [EDGAR])
  - Freebase (already represented as Linked Data)
  - FFIEC XBRL data (FDIC Linked Data Wrapper [FFIEC])

For demo, see <http://xbrl.us/research/appdev/pages/275.aspx#>

## FIOS components



# UC1: Approach – Linking of Data Sources





# UC1: Integrating SEC XBRL and Freebase

vmdeb18.deri.ie:8080/saiku-ui-2.2.RC/

gross profit margin xbrl fios

Unsaved query (1) x

**Cubes**

SEC-Cube-Gross-Profit-Margin

**Dimensionen**

- ▶ Issuer
- ▶ Segment
- ▶ Data set
- ▼ Business business operation
  - industry
    - Business business operation
      - industry root level
    - Dtend
    - Dtstart

**Kennzahlen**

- ▼ Measures
  - Cost of goods sold
  - Sales revenue net

**Spalten**

Cost of goods sold Sales revenue net

**Zeilen**

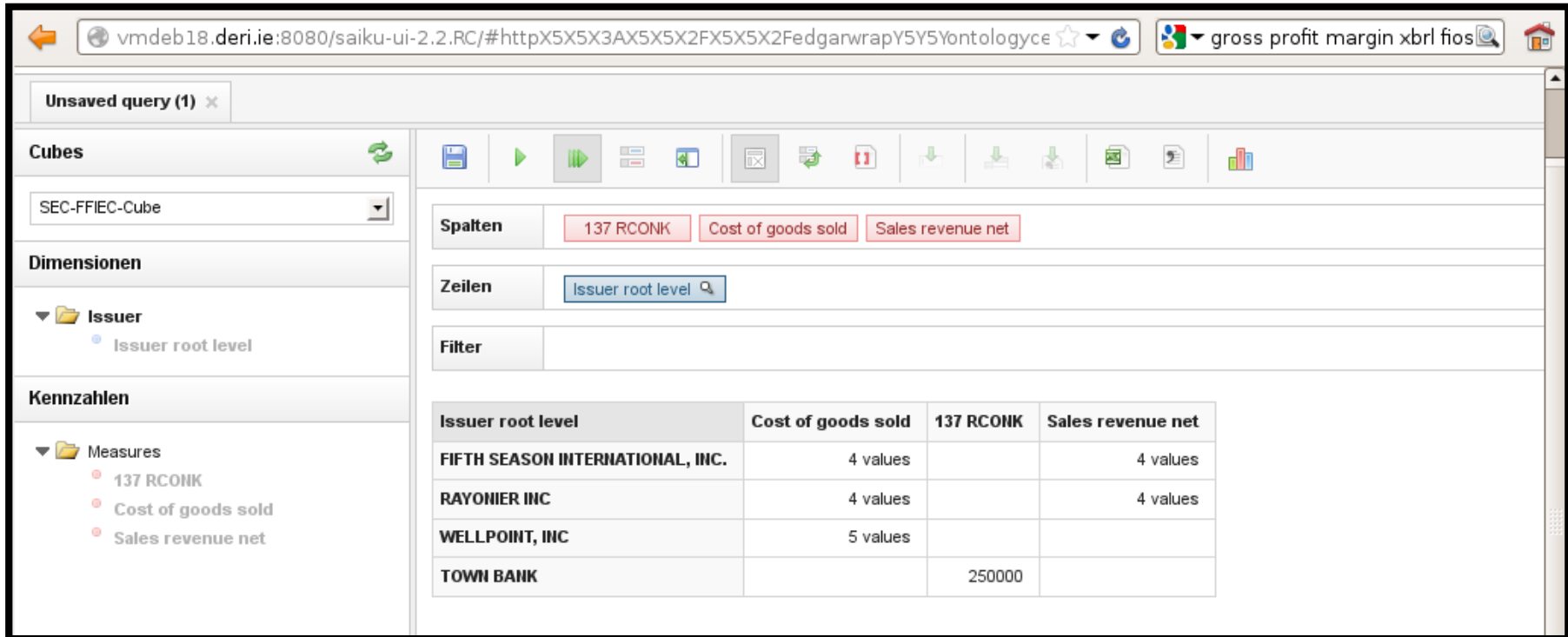
Business business operation industry root level

**Filter**

Business business operation industry root level	Cost of goods sold	Sales revenue net
En financial services concept		24 values
Business sic concept 19876282		24 values
En patent owners and lessors concept		4 values
En meats and meat products concept	4 values	4 values
En meat packing plants concept	4 values	4 values
Business naics concept 2007321	6 values	18 values
En functions related to deposit banking nec concept		18 values
Business sic concept 19876153		4 values
En lessors of real estate concept	12 values	4 values
En fire marine and casualty insurance concept		3 values
En property and casualty insurance concept		3 values
En diversified investments concept		3 values

- Result: Many more Sales revenue net than Cost of goods sold.
- Open problem: Manual effort in creation of additional dimensions.

# UC1: Integrating SEC XBRL with FFIEC XBRL



Unsaved query (1) x

Cubes

SEC-FFIEC-Cube

Dimensionen

▼ Issuer

• Issuer root level

Kennzahlen

▼ Measures

- 137 RCONK
- Cost of goods sold
- Sales revenue net

Spalten

137 RCONK Cost of goods sold Sales revenue net

Zeilen

Issuer root level

Filter

Issuer root level	Cost of goods sold	137 RCONK	Sales revenue net
FIFTH SEASON INTERNATIONAL, INC.	4 values		4 values
RAYONIER IINC	4 values		4 values
WELLPOINT, INC	5 values		
TOWN BANK		250000	

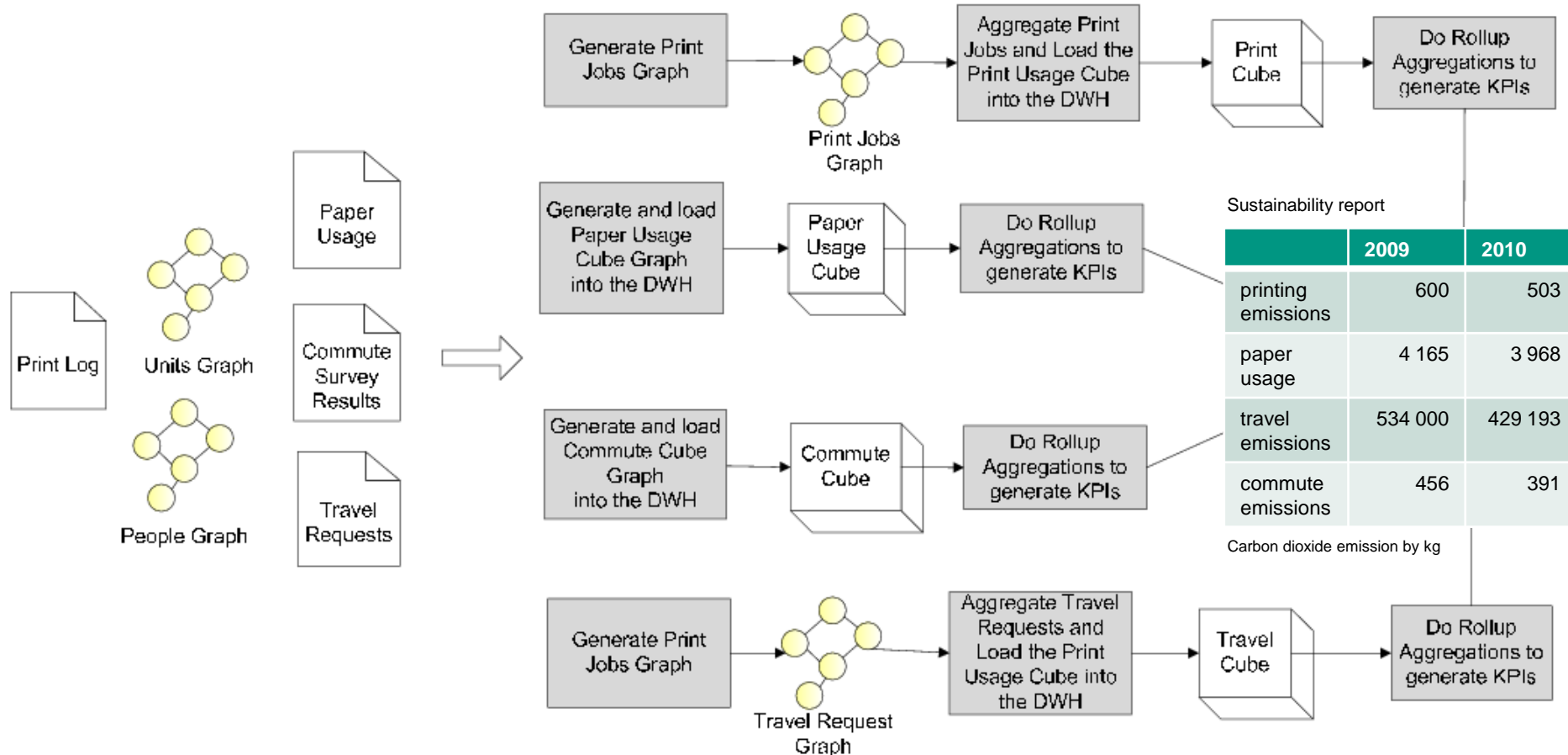
- Result: Selected companies do not provide both SEC and FFIEC information.
- Open problem: Missing links between companies in SEC and FFIEC.

# Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- **UC2: Understanding the Origin and Trustworthiness of XBRL**
- Relation to DDI Lifecycle

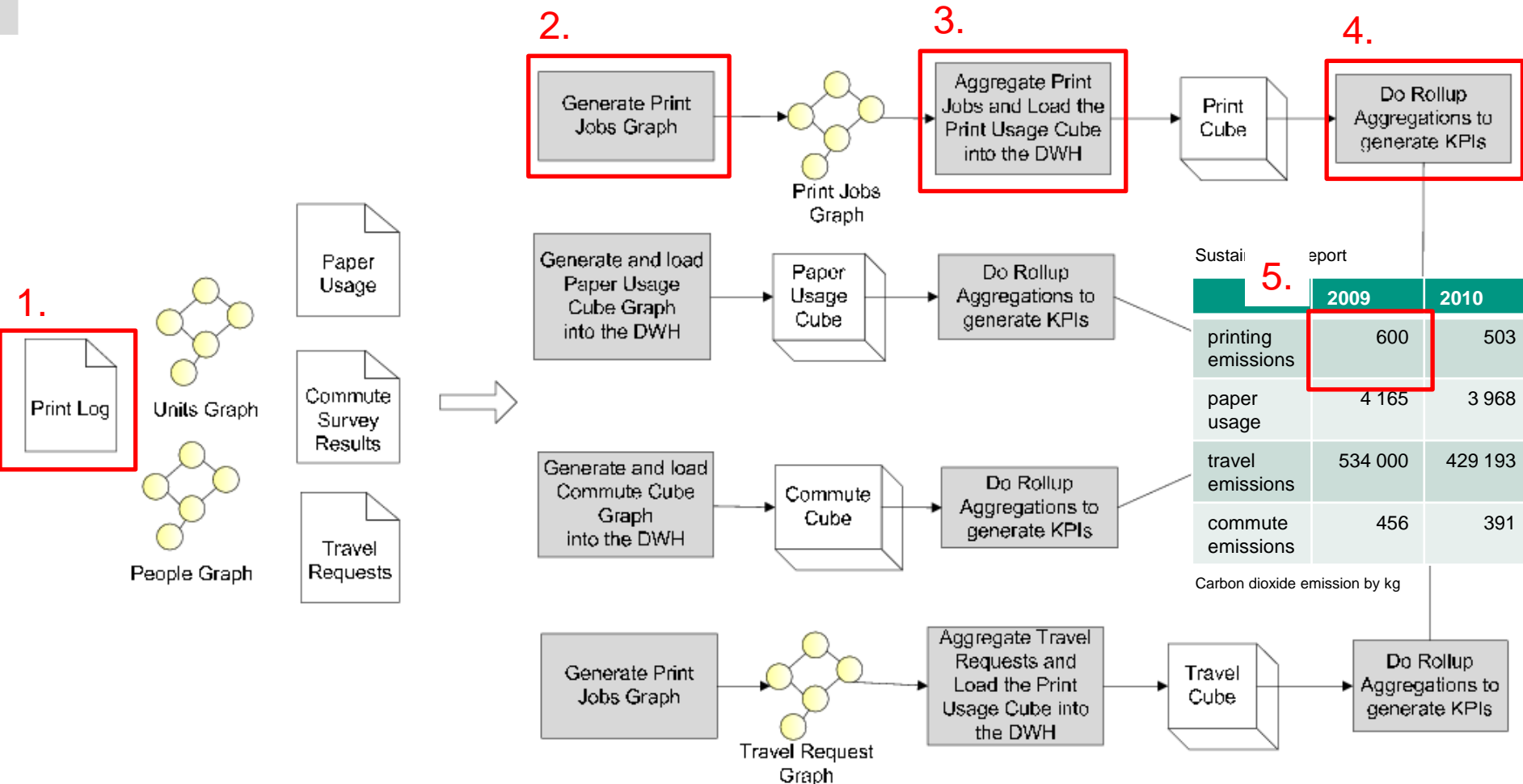
# UC2: Understanding the Origin and Trustworthiness of XBRL

## ■ ETL over heterogeneous data sources



# UC2: Understanding the Origin and Trustworthiness of XBRL

## ETL over heterogeneous data sources



# UC2: Scenario

- How **current** is this information?
- What **aggregation functions** were used?
- What **source** do the raw data come from?

Sustainability report

	2009	2010
printing emissions	600	503
paper usage	4 165	3 968
travel emissions	534 000	429 193
commute emissions	456	391

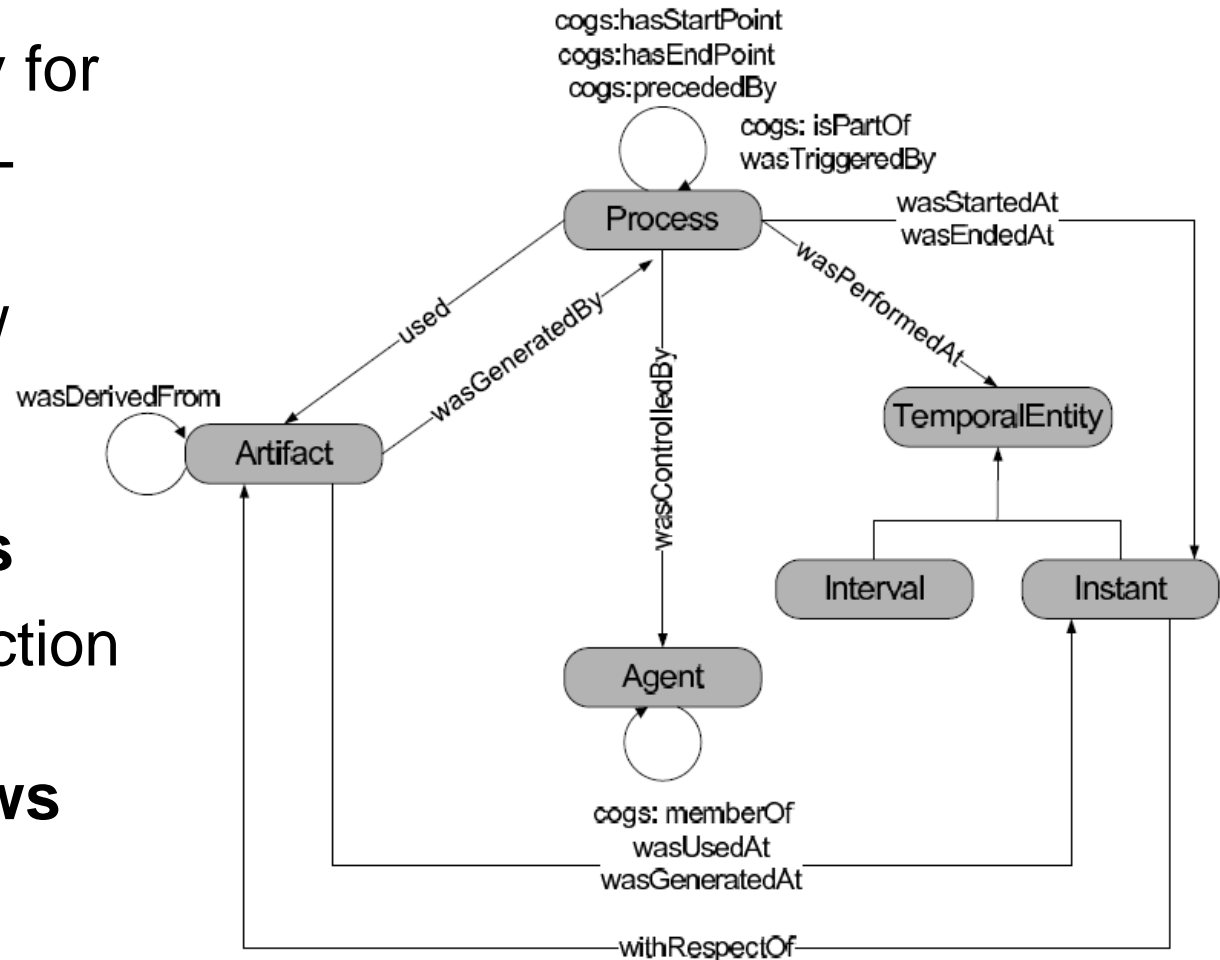
Carbon dioxide emission by kg



## UC2: Approach

# Cogs

- RDF vocabulary for representing ETL elements
- Simple workflow structure of **processes, artifacts, agents**
- Different abstraction levels through **nested workflows**

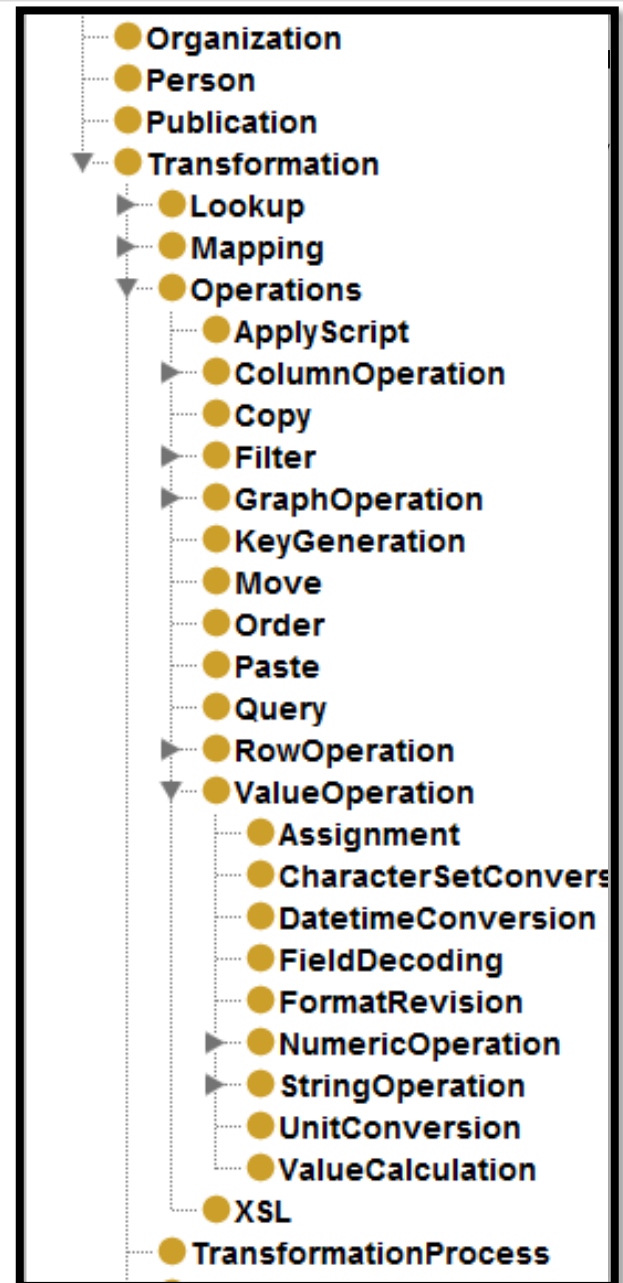


<https://sites.google.com/site/cogsvocab/>

# UC2: Cogs – Structure

- Taxonomy of common ETL concepts, e.g.,
  - Execution
  - State
  - Extraction, Transformation, Loading
  - Layer

Cogs:  
151 concepts  
17 properties



# UC2: Visualising XBRL Report Values



## Report Context

[http://sustainable.deri.ie/resource/report/context/context\\_2010](http://sustainable.deri.ie/resource/report/context/context_2010)

<a href="#">TotalGreenhouseGasEmissionsByWeightResultingFromCommute</a> in kgco2e	44399.86058376993	<a href="#">Detail</a>
<a href="#">AveragePerFTEPaperUsageResultingFromPrinting</a> in sheetPerFTE	269.0551817965995	<a href="#">Detail</a>
<a href="#">AveragePerFTEDistanceResultingFromCommute</a> in kmPerFTE	1675.12573821098	<a href="#">Detail</a>
<a href="#">AveragePerFTEEnergyConsumption</a> in kwhPerFTE	4517.979663268757	<a href="#">Detail</a>
<a href="#">AveragePerFTEGreenhouseGasEmissionsByWeightResultingFromTravel</a> in kgCO2ePerFTE	3784.130755108943	<a href="#">Detail</a>
<a href="#">TotalDistanceResultingFromTravel</a> in km	682896.375	<a href="#">Detail</a>
<a href="#">TotalGreenhouseGasEmissionsByWeightResultingFromEnergyConsumption</a> in kgco2e	266461.2808	<a href="#">Detail</a>
<a href="#">TotalEnergyConsumption</a> in kwh	512425.54	<a href="#">Detail</a>
<a href="#">AveragePerFTEPaperUsage</a> in sheetPerFTE	2120.452678873376	<a href="#">Detail</a>
<a href="#">AveragePerFTEGreenhouseGasEmissionsByWeightResultingFromPaperUsage</a> in kgCO2ePerFTE	34.9874692014107	<a href="#">Detail</a>
<a href="#">TotalGreenhouseGasEmissionsByWeightResultingFromPrinting</a> in kgco2e	503.5122985839844	<a href="#">Detail</a>
<a href="#">TotalDistanceResultingFromCommute</a> in km	189991.3844122141	<a href="#">Detail</a>
<a href="#">TotalGreenhouseGasEmissionsByWeightResultingFromPaperUsage</a> in kgco2e	3968.25	<a href="#">Detail</a>
<a href="#">TotalGreenhouseGasEmissionsByWeightResultingFromTravel</a> in kgco2e	429193	<a href="#">Detail</a>
<a href="#">AveragePerFTEGreenhouseGasEmissionsByWeightResultingFromPrinting</a> in kgCO2ePerFTE	4.439392941281084	<a href="#">Detail</a>
<a href="#">AveragePerFTEGreenhouseGasEmissionsByWeightResultingFromEnergyConsumption</a> in kgCO2ePerFTE	2349.349424899754	<a href="#">Detail</a>
<a href="#">TotalPaperUsage</a> in sheet	240500	<a href="#">Detail</a>
<a href="#">AveragePerFTEGreenhouseGasEmissionsByWeightResultingFromCommute</a> in kgCO2ePerFTE	391.4669576567956	<a href="#">Detail</a>

# UC2: Visualisation of Provenance Descriptors



### KPI Name

[TotalGreenhouseGasEmissionsByWeightResultingFromPrintin](#)

### Context URI

[http://sustainable.deri.ie/resource/report/context/context\\_2010](http://sustainable.deri.ie/resource/report/context/context_2010)

### Unit

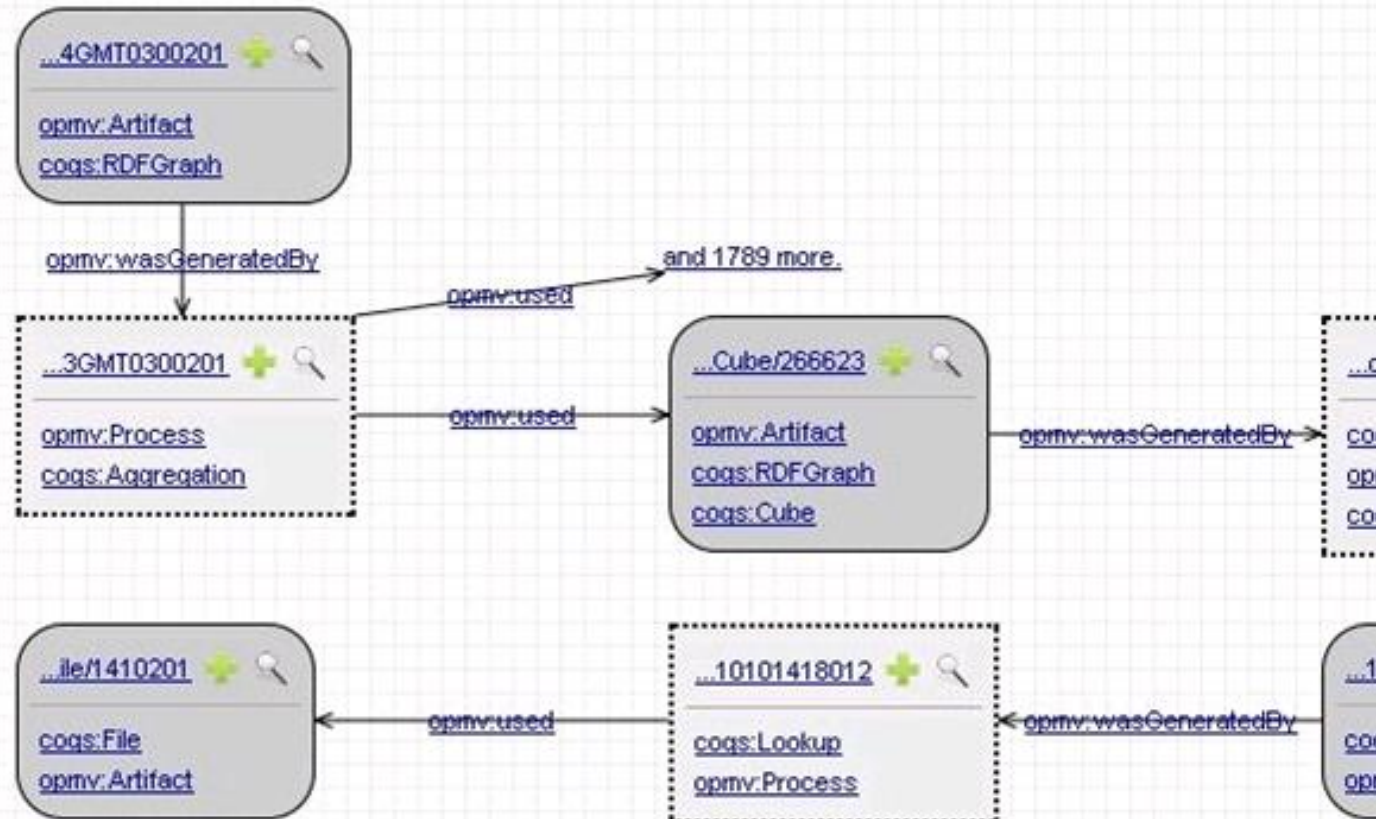
<http://sustainable.deri.ie/measurementunits#kgco2e>

### Value

503.5122985839844

### GRI KPI Compliance

EN16 - Total direct and indirect greenhouse gas emissions by weight.





# UC2: Visualisation of Provenance Descriptors



### KPI Name

TotalGreenhouseGasEmissionsByWeightResultingFromPrintin

### Context URI

http://sustainable.deri.ie/resource/report/context/context\_2010

### Unit

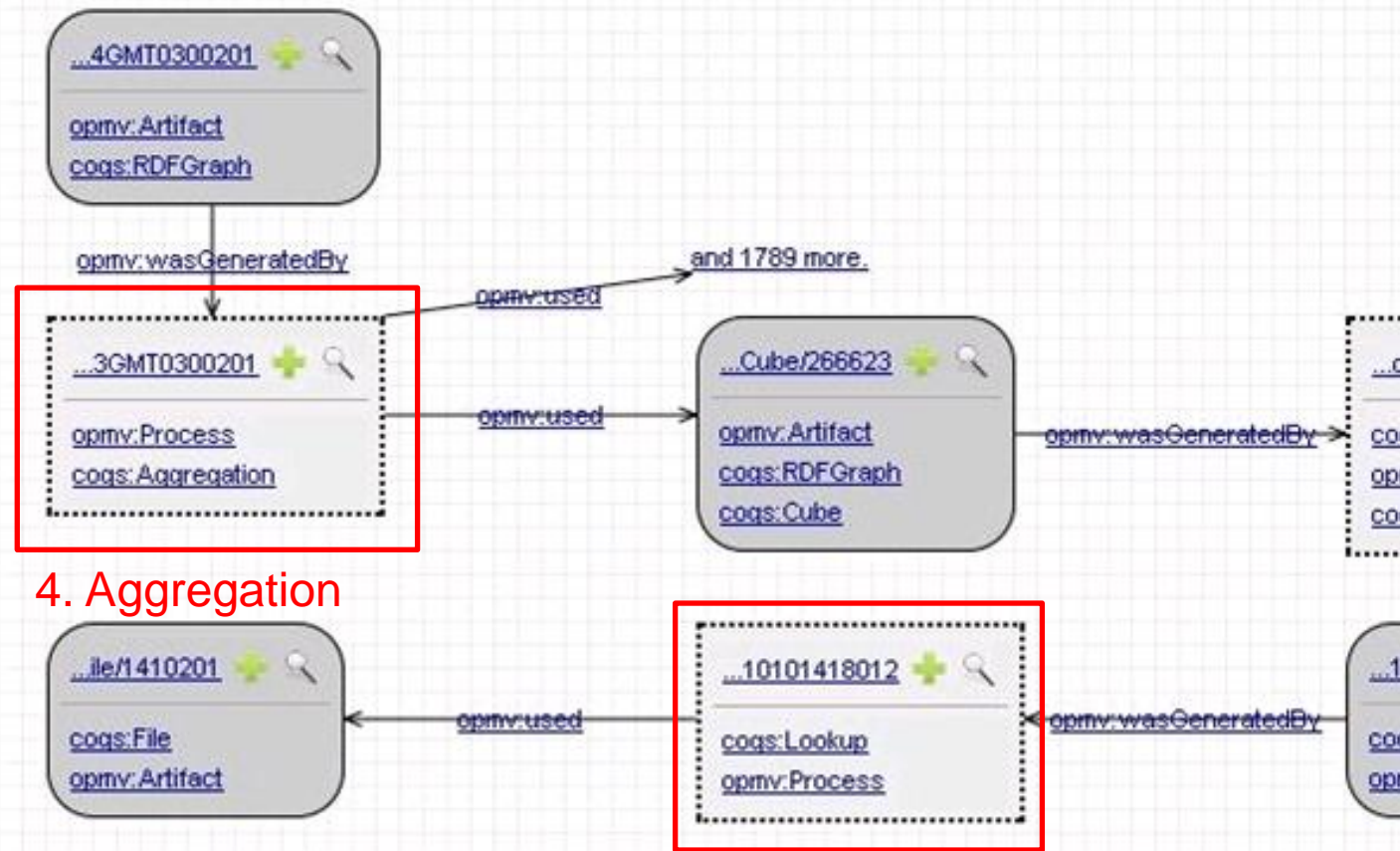
http://sustainable.deri.ie/measurementunits#kgco2e

### Value

503.5122985839844

### GRI KPI Compliance

EN16 - Total direct and indirect greenhouse gas emissions by weight.



4. Aggregation

1. Lookup

# UC2: Visualisation of Provenance Descriptors

### KPI Name

[TotalGreenhouseGasEmissionsByWeightResultingFromPrintin](#)

### Context URI

[http://sustainable.deri.ie/resource/report/context/context\\_2010](http://sustainable.deri.ie/resource/report/context/context_2010)

### Unit

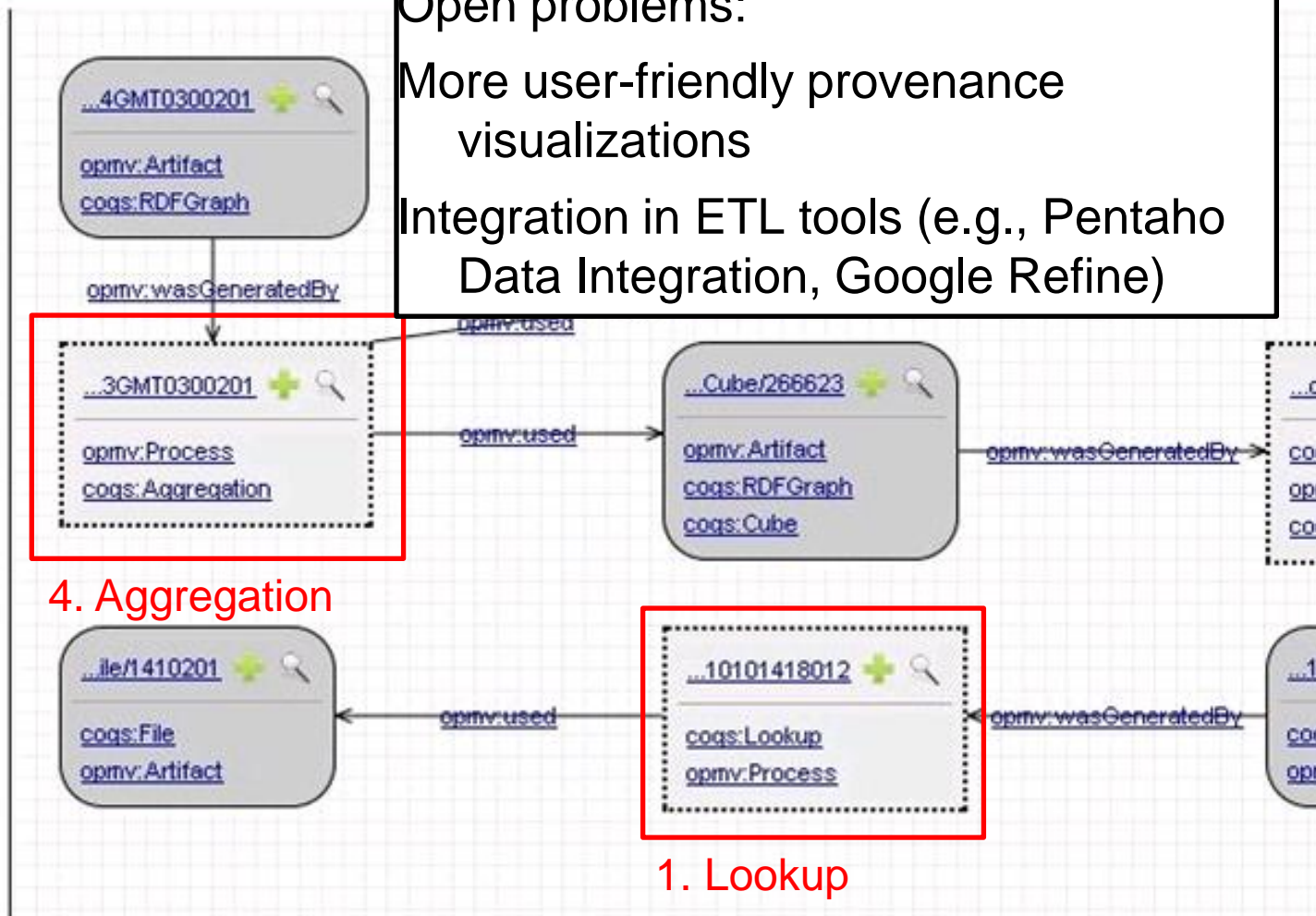
<http://sustainable.deri.ie/measurementunits#kgco2e>

### Value

503.5122985839844

### GRI KPI Compliance

EN16 - Total direct and indirect greenhouse gas emissions by weight.



Open problems:

More user-friendly provenance visualizations

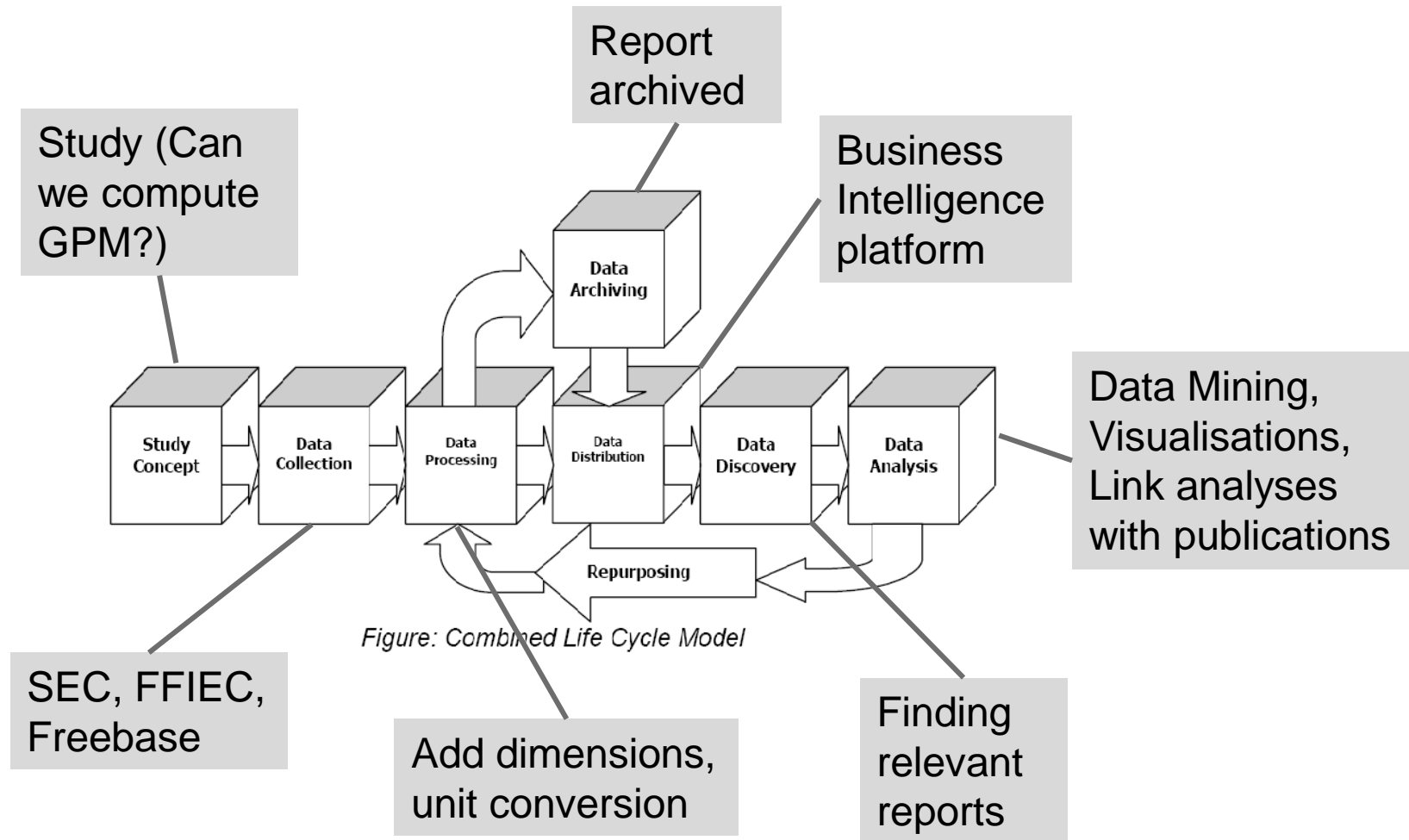
Integration in ETL tools (e.g., Pentaho Data Integration, Google Refine)



# Outline

- Extensible Business Reporting Language (XBRL)
- UC1: Integrating XBRL with External Data Sources
- UC2: Understanding the Origin and Trustworthiness of XBRL
- **Relation to DDI Lifecycle**

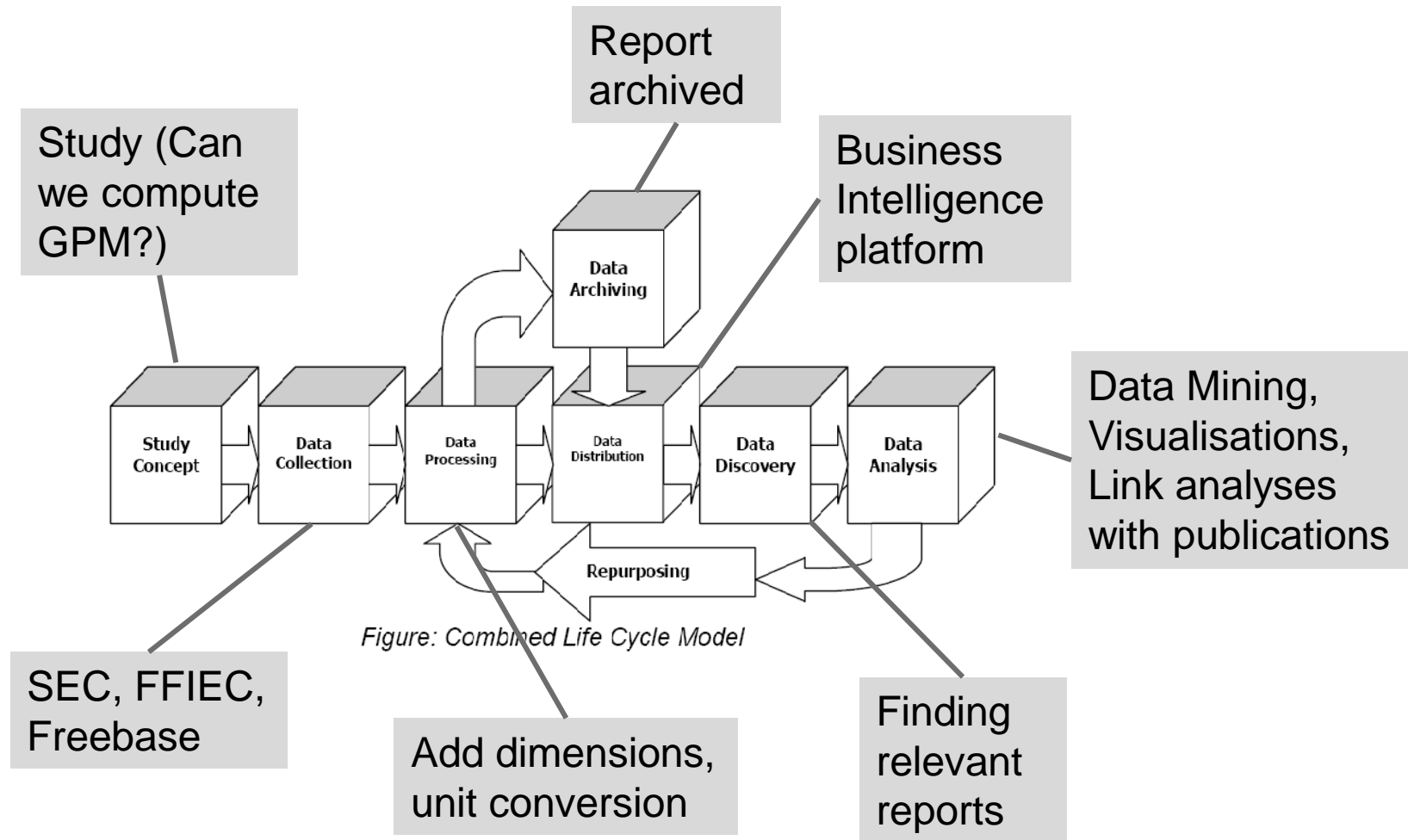
# Relation to DDI Lifecycle



# Conclusions

- XBRL interesting data source to be published as Linked Data for integration with other finance data.
- Open challenges in „mechanics“ and „analysis“ of Financial Linked Data.
  - Preprocessing, matching, and understanding of data
- DDI interesting for XBRL use cases

# Thank you for your attention, questions, and feedback

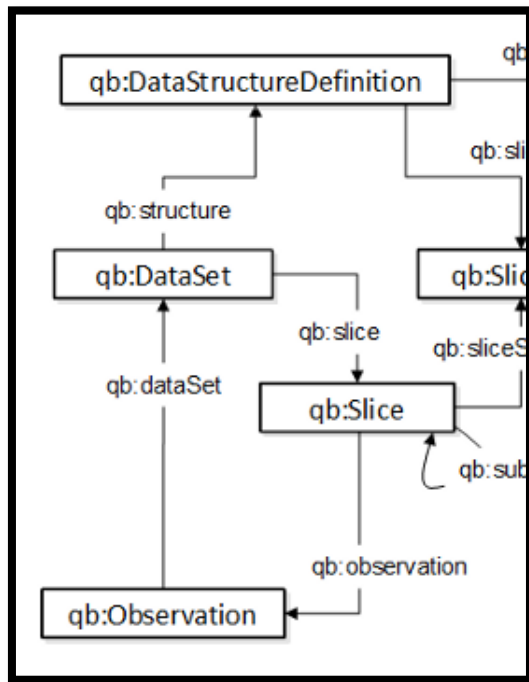


# References

- [XBRL SPEC] <http://www.xbrl.org/Specification/XBRL-RECOMMENDATION-2003-12-31+Corrected-Errata-2008-07-02.htm>
- [SEC PCY] <http://www.sec.gov/spotlight/xbrl/financial-statements.shtml>
- [XBRL example] <http://www.sec.gov/Archives/edgar/data/52827/000119312510238973/0001193125-10-238973-index.htm>
- [DEBRECENY] [http://web.ku.edu/~eycarat/myssi/\\_pdf/2-Debreceeny-XBRL%20Ratios%2020101213.pdf](http://web.ku.edu/~eycarat/myssi/_pdf/2-Debreceeny-XBRL%20Ratios%2020101213.pdf)
- [XBRL AM] <http://xbrl.org/Specification/abstractmodel-primary/PWD-2011-10-19/abstractmodel-primary-PWD-2011-10-19.html>
- [EDGAR] <http://edgarwrap.ontologycentral.com/>
- [FIOS] <http://xbrl.us/research/appdev/pages/275.aspx#>
- [EDGAR RAY] <http://edgarwrap.ontologycentral.com/archive/52827/0001193125-10-238973#ds>

# Backup: UC1: Why is QB such a good fit?

## ■ Using RDF Data Cube Vocabulary (QB) for representing XBRL data

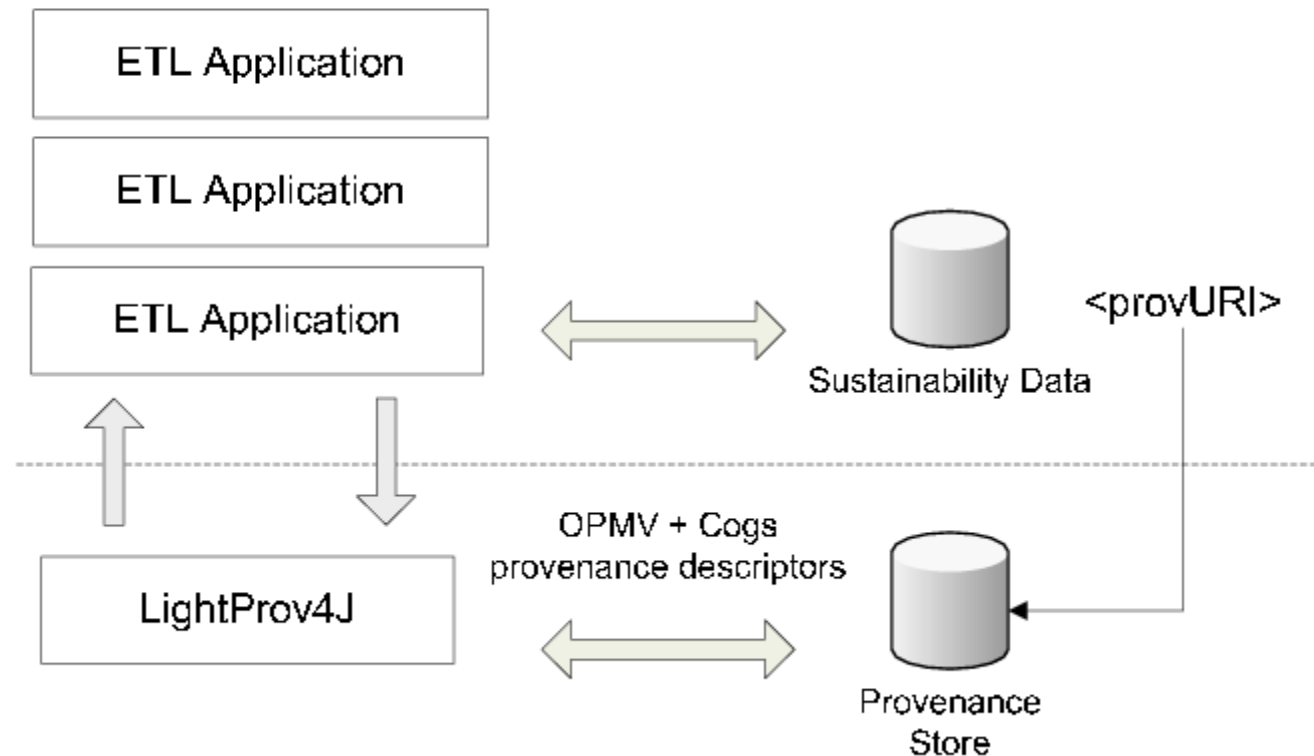


- Aggregate Data: 1) avg, count, sum 2) XBRL formulas/calculation arcs
- Numeric: financial facts
- Time series: valid start and end date of facts
- Multidimensional: filing, issuer, segment...
- Highly structured: concepts from taxonomies

<http://www.w3.org/TR/vocab-data-cube/>



# UC2: Approach – Architecture with Provenance-aware ETL Applications



# Backup: Future Work Summary

- Further reduce “mechanics” needed
  - Creation of additional dimensions
  - Filtering for certain companies
  - Identification of identical companies in SEC and FFIEC
- Support more expressive “analyses”
  - Concept hierarchies
  - Formulas/business rules for missing values
  - Segment information (e.g., salary of CEO)

# Relation to DDI Lifecycle and Use Cases

- “DDI For Use within a Research Project”
- Describe analysis goals and results
- Publish results of analyses (including queries) again as Linked Data and allow other users to enhance it, e.g., add annotations, add additional information to dimensions.

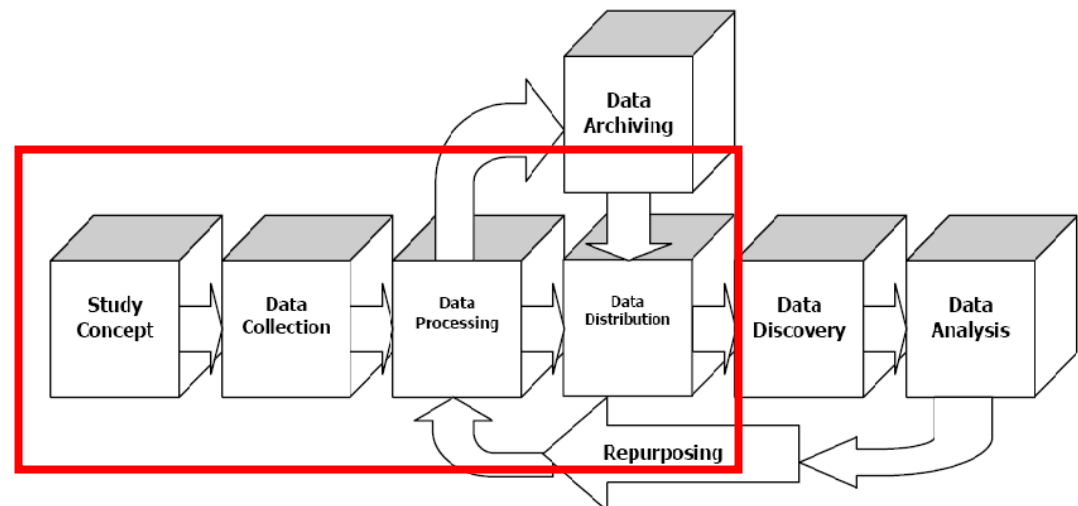


Figure: Combined Life Cycle Model

# Relation to DDI Lifecycle and Use Cases

## (2)

- “Data Dissemination/Data Discovery”
- “Finding and Linking Publications related to Data”
- Link analyses with publications
- Many data sources are finance-relevant, the question is to find ‘‘...’’

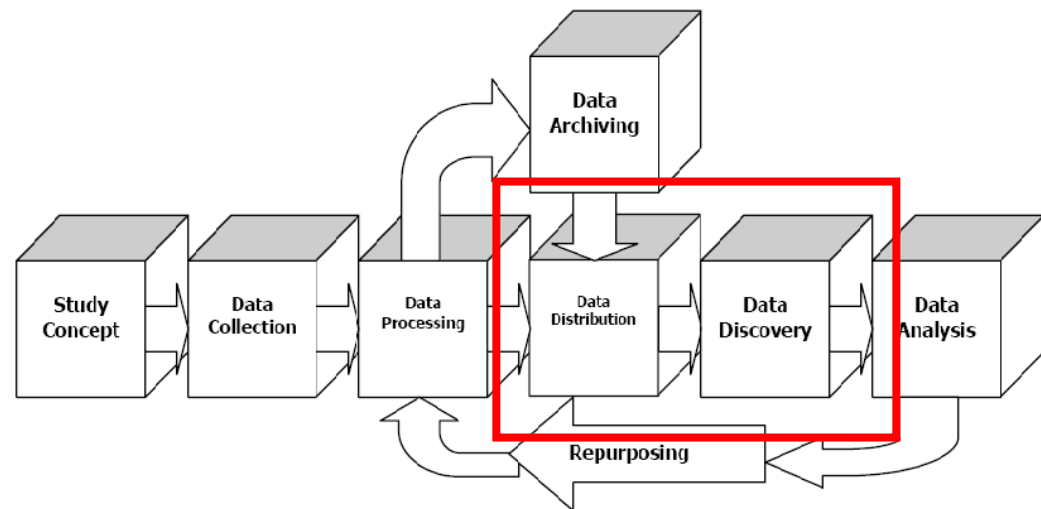


Figure: Combined Life Cycle Model

# Relation to DDI Lifecycle and Use Cases

## (3)

- “Metadata Mining for Comparison, etc.”
- “Links to external thesauri”
- Linking comparable financial information
- Publishing taxonomy information and extensions
- Integration of “Web Tables”

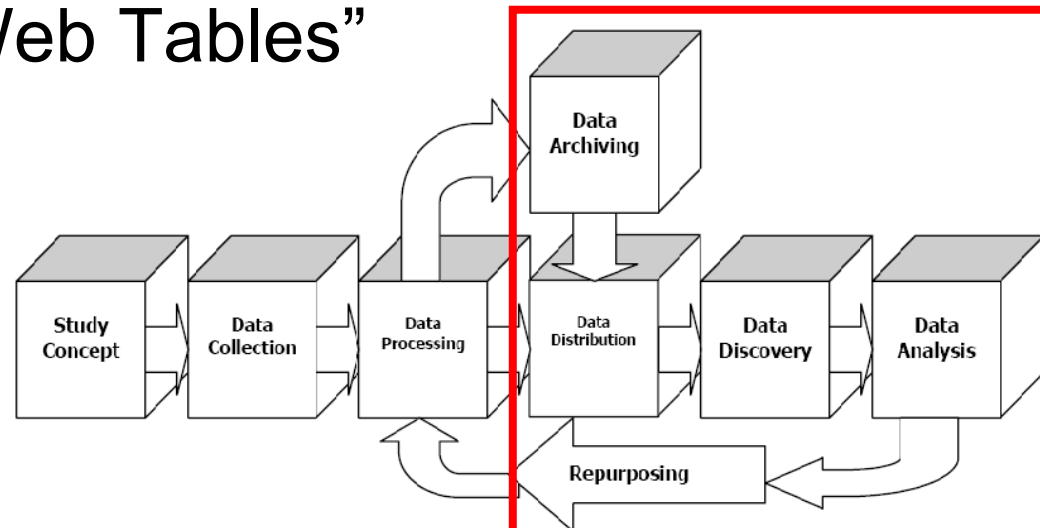


Figure: Combined Life Cycle Model